Effects of Subsurface Heterogeneity on Groundwater Flow and Transport: A DoD HPC Challenge Project

J. F. Peters, S. E. Howington, F. T. Tracy, J.P. Holland,
U.S. Army Waterways Experiment Station
Vicksburg, MS 39180

R. S. Maier
Army High Performance Computing Research Center,
Minneapolis, MN 55415

Abstract

The Department of Defense has thousands of contaminated subsurface sites that require some level of restoration. Costs of these cleanups are in the billions of dollars. Cleanup of these sites is hampered by heterogeneity inherent in porous media. High performance computing (HPC) tools are used to create a simulated laboratory to determine laws relating the evolution of micro-scale quantities of heterogeneous porous media, which are represented statistically, to the mean fields computed from numerical analyses of moderate resolution. This effort is based on the philosophy that considerable spatial detail can be sacrificed provided an appropriate equivalent medium can be defined. An equivalent medium is appropriate if it alters the statistics of a contaminant field in the same manner as the actual medium. The paper discusses the need for the research, the numerical procedures under development, and the difficulties encountered in creating sufficiently high resolution HPC simulations to supply data for that development.

1. Background

The Department of Defense (DoD) has thousands of contaminated groundwater sites, the result of past military activities, whose remediation is estimated in the tens of billions of dollars. Groundwater flow and transport modeling is a critical element in predicting the effectiveness of groundwater remediation schemes. High Performance Computing (HPC) is a key enabling tool to interconnect the many, varied simulation capabilities required to advance the science of groundwater remediation and to provide for more cost effective cleanup actions. The US Army Engineer Waterways Experiment Station is engaged in research and development that provides a seamless link to tools that combine diverse field data into accurate site characterizations, and perform computer simulations of remediation processes (Holland, 1996). However, the simulation models on which this capability is founded, are still not adequate to simulate real, heterogeneous media that make up the natural groundwater environment. The inadequacy stems from the multi-scale structure of natural media that is masked by the continuum formulations upon which numerical models are based. The range of scales in natural subsurface media is simply too great to be spanned by any present, or foreseeable, HPC resource. To advance the science of multi-scale simulation, HPC may be viewed as an extension of the laboratory. By stretching present resources to perform scale-spanning simulations, stochastic models are being created from data that could not be obtained previously from physical experiments. A detailed description of this process is presented for dispersive transport, the focus of this year's Challenge project investigations.

1.1 Flow and Transport

While chemical and biological processes are the focus of many remediation designs, prediction of these processes depends on an accurate depiction of fluid flow and chemical transport. For example, in a heterogeneous medium the effect of differential advection on conservative transport is incorrectly characterized in traditional field-scale subsurface models as being caused by other processes. Under such conditions, the numerical coefficients that characterize chemical or biological processes appear to vary with the scale of the problem and processes that are occurring. Existing numerical models depict this growth rate as a scale effect that causes dispersion parameters obtained from laboratory experiments to be orders of magnitude different from values observed in the field. Similarly, parameters that model early growth rates of plumes do not predict the growth of mature plumes in traditional models. Because these parameters are supposed to be constant, the scale effect is spurious and reflects the inadequacy of traditional continuum-based relationships used in existing production groundwater codes. This inadequacy can lead both to over-design of remediation systems (through over-estimation of contaminant plume spreading and strength), or under-design (due to excessive dilution of the contaminant plume based on model parameter selection). Both of these scenarios (increased costs due to over-design and system failure resulting in increased human and environmental risk) are unacceptable.

1.2 High Resolution Discrete Models

Discrete models offer the ability to simulate macroscopic behavior by providing a bridge between fundamental-scale models and engineering-scale continuum models. Spatial detail at the fundamental scale requires data that are typically not assessable. Thus, research efforts are also directed at developing methods to devise statistically equivalent site properties (see Poeter et al., 1997). Examples of discrete medium modeling are highly-resolved continuum models (e.g., the PARFLOW model (Tompson et al. 1997), developed by DOE's Lawrence Livermore National Laboratory and implemented within the DoD Groundwater Modeling System – the GMS), macroscopic discrete network models, and particle-based transport models. In particular, the statistical network model (SNC) of Peters and Howington (1997) and Howington et al. (1997) has provided a virtual laboratory to understand the relationship between micro-scale quantities versus those usually associated with continuum models. The particle-based transport model provides a convenient means to implement physics gleaned from high-resolution discrete models in production-scale, engineering numerical codes of much lesser resolution. For example the ADH code of Schmidt (1997) provides a multi-resolution capability to optimize the relationship between fidelity of simulation and HPC resources.

1.3 The Continuum

A continuum is a mathematical construct. A domain is said to be a continuum if a material point occupies every point within its boundaries. All media are, at some level, composed of particles and therefore none satisfies the mathematical definition of a continuum. The applicability of the continuum concept is justified by the notion that the physical size of discrete components is small compared to the aperture size for observing the medium. Therefore, the practical definition of the continuum involves some averaging scheme in which a property is averaged over some suitable, finite representative elemental volume (REV) in the actual medium. This REV value is then assigned to each infinitesimal differential element in the continuum. For this procedure to be valid, the medium is either sufficiently fine grained to allow a single property to be defined or there is such a large separation between scales that the medium can be broken into facies, each having its own property. In each case, there exists a scale separation between the scale of averaging and the scales of material variability. The REV approach assumes that the properties are distributed such that an aperture size can be found for which the spatial average is constant. That is, the medium is statistically asymptotic. Two difficulties arise from application of the REV. First, most natural porous media are not statistically asymptotic. Features may exist over a range of scales such that no single aperture size is suitable to define a REV. Second, the physical and chemical phenomena of interest have natural scales that interact with those of the medium. For chemical transport, models that are artificially diffusive exaggerate their ability to support key chemical reactions.

To use knowledge gained from laboratory experimentation and pore-scale numerical analyses in a field-scale engineering code (e.g., ADH), one must embed the statistics of the small-scale simulation in the material description for the field scale. Thus, the mathematical description for the field-scale, homogeneous material should have the same dispersive properties as the heterogeneous material used in the direct simulation. The classical advection-dispersion equation (ADE) attempts this through a Fickian description of dispersion. Unfortunately, the Fickian model lacks a correlation structure found in real media. Non-local models of dispersion (Koch and Brady,1987; Neuman, 1993; and Cushman and Hu, 1995) provide correlation structure but, to date, have not provided versatile, practical transport models.

1.5 Searching for the Equivalent Medium

If remediation sites could be modeled entirely with a pore-scale model, there would be no need to parameterize many effects at the pore scale and larger. However, because modeling at the pore scale is technically infeasible for engineering applications, one is left to determine how much detail will be resolved and modeled directly, and how much will be accounted for through the averaging formulations. Thus it is necessary to replace a spatially resolved media with one that is equivalent in some sense. For example, consider a cubical element containing a heterogeneous mixture of materials. For a flow simulation it is generally sufficient to replace the mixture with an equivalent medium that has the same hydraulic conductivity which relates the flow rate to the mean gradient. Note that in this case the partial differential equation that describes the highly resolved system is the same one that describes the equivalent one. However, there are physical effects not captured by the equivalent media. In some cases, these effects may not be captured by any equivalent property unless the governing partial differential equation is also modified. That is, the equivalence is established by the equation, not merely its parameters. For example, when a heterogeneous media is homogenized, the large-scale governing equation must contain a dispersive term even though the micro-scale equation predicts no dispersion. The dispersion comes not from the governing equation but from interactions among the small-scale elements.

An equivalent flow equation that considers only the mean flow rate misses several factors essential to transport. First, heterogeneous features have sizes that are not accounted for in such an equivalent medium. Thus, the equivalent medium can support singularities at boundaries and high gradients that would not be present in the heterogeneous medium. Therefore the equivalence is valid away from boundaries. Second, if we wish to couple the flow equation to the transport equation, we must recognize that we need more that the mean velocity to drive transport. Thus our equivalent medium must contain a statistical description of flow velocities that is coupled to the transport equation. The velocity statistics are traditionally parameterized by the dispersivity coefficient that depicts the velocity as a white noise characterized by a mean and variance. This description of velocity ignores the size of heterogeneous features and is only valid as a asymptotic case when the dispersing plume is large compared to the size of the feature (Peters and Howington, 1997; Tompson and Gelhar, 1990). The effect of ignoring size is that the dispersion rate will appear to change as the plume grows. An improved statistical description considers a correlation among velocities in either time and/or space. This description gives an improved prediction of dispersion without having spatial information on the distribution of the heterogeneities. In this case, ***the basis of equivalence is that the media alters the statistical descriptors of the plume (centroid, size, skewness, etc.) in the same manner as the real media.***

1.6 The Approach

This investigation's approach is similar to that taken on soil deformation described by Horner (1997), Horner et al. (1994 and 1997) and Hryciw et al. (1997), in which discrete element models (DEMs) of particle systems were used to develop evolution laws for interparticle forces. These forces fed macro-level constitutive relationship for large deformations. The DEM provided a valuable laboratory to understand the equivalent medium problem at a fundamental level. Once the forms of the statistical relationship were established, parameters to model a particular material could be obtained from standard laboratory experiment. In this research a number of numerical simulation codes that model porous at various scales are used in concert to devise statistical laws. In the following section, the particular example of how the replacement of spatial detail with its statistical equivalence has been accomplished.

2. The Network and the Continuum

To demonstrate the method developed in this investigation, the statistical network code (SNC) was used to create a detailed medium whose transport properties would be replicated by a very low- resolution model. The SNC was ideal because its ability to capture the multi-scale dispersive nature of real media has been demonstrated with computational experiments (see Peters and Howington, 1997 and Howington et al., 1997). The SNC simulates multi-scale media by depicting porous media as a dense network of flow paths (called throats), along which contaminants are transported by one-dimensional advection. Dispersion occurs as a result of mixing at network connections and different travel times between connections. In particular, the SNC simulations displayed the same growth rate of contaminant plumes that were observed in field experiments using parameters that are constant. The realistic growth rates emerge from the simple interaction of network quantities. Thus, large-scale flow networks obviate apparent scale effects that plague existing numerical models used for groundwater remediation.

The network is an analogy to a porous medium. To relate this discrete network to the continuous representation of the porous medium, a filter is employed. A variation on the REV concept is used to define a medium property as a weighted spatial average. A discrete property such as mass can be expressed as a continuous density via a convolution integral of the domain.

$$? (x_i) ? \int_{?}^{?} ? (x_i ? x_i?) m(x_i?) dx_i?. \qquad (1)$$

This definition is advantageous from two viewpoints. From an operational standpoint, gradients of the quantity can be defined in a straightforward fashion.

$$\frac{? ?}{? x_i} ? \int_{?}^{?} \frac{? ?}{? x_i} (x_i ? x_i?) m (x_i?) d x_i?. \qquad (2)$$

From a conceptual viewpoint, the effect of the average is to filter information such that the continuum idealization is a coarse-grained version of the original medium. Each scale is still represented, but the contribution of finer scales is reduced. In recent work by Horner (1997), a discrete element model of granular media was approximated as a coarse-grained particulate using this smoothing concept. The method amounted to approximating the difference equations of the discrete element model with a weighted-residual procedure whereby the kernel function of Equation 1 becomes a test function.

The effect of a filter is to create a continuous space onto which quantities of the network solution and the porous medium can be mapped (Figure 1). The fine-grained detail in the network and the porous medium differs greatly but, because larger-scale structures are the same for both representations, they appear similar when viewed through the filter. This fact is most apparent when the pressure solution is considered. The network and porous media are both discrete linear systems. It can be shown formally that when the degrees of freedom of linear networks are reduced by averaging, an infinity of these networks contain identical averaged solutions. It is this equivalence property that allows conductivity measurements to be made on two specimens that are profoundly different at the micro-scale, yet give rise to the same macro-scale conductivity. The case for non-linear behavior is not as clear-cut. However, in practice properties of a medium are prescribed routinely on the basis of a finite number of specimens even without proof for a formal equivalence property.

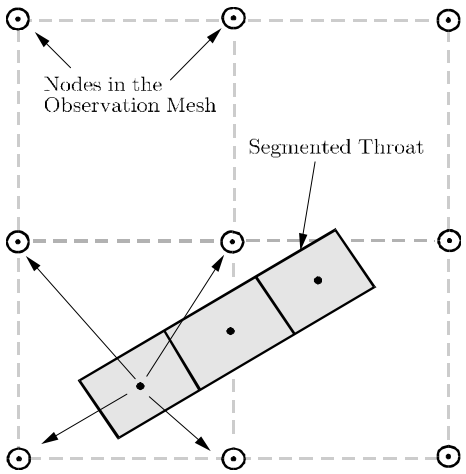
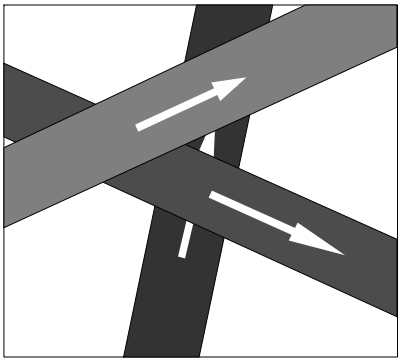


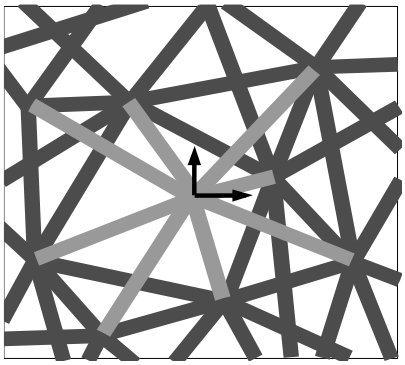
Figure 1. Schematic showing a simple spatial averaging filter applied to an individual throat in a discrete network.

2.1 The Inverse Problem

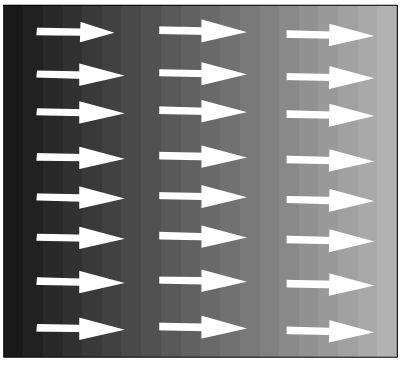
Specification of properties for individual throats in the network requires an inversion of the filtering procedure. The equivalence property that justifies filtering also implies that the inverse problem is ill-posed, because the fine-scale network that produces a particular filtered solution is not unique. Therefore, the goal of inversion is to find a statistical distribution of network properties that gives rise to the desired averaged medium property. To simulate flow and transport, we must create a network that displays the desired, macroscopic conductivity and apparent dispersivity.



Throat Scale



Patch or Darcy Scale



Macroscopic Scale

Figure 2. The three, distinct scales associated with a discrete network.

Conductivity. The conductivity of a network is an equivalence continuum (average) property measured away from boundaries. The equivalence requires two levels of averaging. First, discrete throat conductivities must be averaged to a local conductivity value at the patch scale (Figure 2). Second, these local values must be appropriately averaged to an equivalent medium property at the domain scale (Figure 2). This second step is more straightforward since the following results of Gelhar (1993) can be applied directly to relate the effective hydraulic conductivity (K^e) to the geometric mean (K^g), the problem dimension (n), and the patch-scale conductivity variance (σ_f^2):

$$K^{e}_{ij} = K^{g}_{ij}\left[1 - \left(\frac{1}{2} - \frac{1}{n}\right)\sigma^{2}_{f}\right] \qquad (3)$$

Patch-scale conductivities are computed as an average based on throat conductivities. This patch-average conductivity would be observed if the patch were removed from the network and placed in a constant, unit-gradient total head field.

$$K_{ij} = \frac{1}{2}\sum_{t=1}^{C_N} n_i \frac{Kt}{L_t}\left(H^{c}_{j} - n_j L_t\right) \qquad (4)$$

Where, H^c_j is computed to enforce conservation of mass at the center node.

$$H^{c}_{j} = \sum_{t=1}^{C_N} K_t n_j \Big/ \sum_{t=1}^{C_N} \frac{K_t}{L_t}. \qquad (5)$$

Dispersion. Dispersion is governed by the variance in the velocity from the mean value. This variance arises from two sources. There is an essential variance caused by the various flow path directions imposed by the network topology. Even a "homogeneous" network in which all throats have the same conductive capacity will produce variations in flow-path direction. There is also variance caused by variability in throat properties. The growth in dispersivity depends on the spatial correlation among velocities (Peters and Howington, 1997) which is controlled by throat length. It was found that for all networks, the dispersivity grows from a base value, determined by the fineness of the network, to an asymptotic value that depends on the global velocity variance. Howington et al. (1997) showed that a target asymptotic value of (1) lateral and (2) longitudinal dispersivity could be estimated from a mapping to (1) variance in throat conductivity and (2) the distribution of throat lengths in the network.

3. Breaking Out of the Box

The limitation of the network comes from the level of resolution that can modeled by the network model. The computational requirements of the network increases exponentially with the range of scales represented (as measured by smallest network component relative to domain size). While very realistic contaminant plumes can be created, contaminant concentrations are only accurate when results are averaged at a scale considerably greater than the smallest network component. The key to extending the range of the computation lies in the observation that the dense network is required to create the variable (random) flow needed to drive dispersion. Peters and Howington (1997) noted that the averaged flow field is smooth, even for highly heterogeneous networks, and can be reproduced by a traditional flow model using less computational resources. The implication of this observation is that a pressure solution is not needed on a highly-resolved heterogeneous medium provided the velocity statistics that drive the dispersive transport can be replicated by other means. If the velocity statistics can be modeled as a random process that is driven by the average flow field, the high-resolution flow computations can be avoided. The classical advection-dispersion equation (ADE) attempts this through a Fickian description of dispersion. Unfortunately, the Fickian model lacks a correlation structure found in real media. Non-local models of dispersion (Koch and Brady, 1987; Neuman, 1993; and Cushman and Hu, 1995) provide correlation structure but, to date, have not provided versatile, practical transport models.

A promising approach for bridging the model scales for contaminant transport and remediation processes is correlated particle tracking (Scheibe and Cole, 1994). In correlated particle tracking, a highly-resolved numerical simulation provides statistical velocity distributions used to drive a particle-based transport scheme. In effect, the particle tracking approach directly simulates the behavior that non-local models attempt to capture more compactly by their integral formulations. The History-Dependent Dispersive Transport (HD2T) model is a particle-based transport code designed by this project's investigators to exploit the insensitivity of the mean flow field to the degree of heterogeneity. Dispersion is controlled primarily by deviations from mean flow, which necessarily display a history dependence that can be expressed through a velocity correlation structure. It follows that an equivalent media can be created that has the same dispersive effects as the actual heterogeneous medium provided temporal/spatial correlations are included in velocity statistics.

In HD2T, the statistical components of the particle tracking velocities are time-correlated such that they replicate those observed in the network code. As for the traditional random walk model, the advection velocity v_i consists of a mean flow v^m_i and a random deviation from the mean v^d_i.

$$v_i = v^m_i + v^d_i. \qquad (6)$$

By definition, the mean of v^d_i is zero. If v^d_i has a Gaussian distribution, Fick's law can describe the dispersion created by the random component, resulting in the traditional ADE. However, in a media having a finite length scale, $v^d_i(t)$ is correlated with $v^d_i(t-?)$, for all $0<? <t$. This correlation is readily seen for transport in a network because once a "particle" of mass enters an advection path, its velocity remains constant until it exits the path. If the plume of particles is small relative to the sizes of the advection paths, the advection velocities display high correlation. This correlation persists as long as a significant portion of particles retain the same velocities over a time that is long compared to the time required for a plume to travel a distance comparable to its own dimension. If the pathways are short relative to the plume size, the time required for particles to enter and exit pathways is short compared to the time required for the plume to move a significant distance, the correlation effect is lost. Once correlation is lost, a diffusion model for dispersion is appropriate. Hence, as a plume grows, an asymptotic state is obtained for which the traditional ADE approximates the growth rate.

A potential problem with this approach is that the dispersivity in a correlated system is time dependent, requiring the history of concentrations to be maintained. For long simulations, any HPC resource would be swamped by the required calculations. The alternative to a memory-based system is one based a set of state variable that control dispersive flux that evolve in time. Preliminary results showed that a random walk procedure could generate the ensemble flow statistics to mimic dispersion in the network.

The network itself suggests a mathematically unsophisticated procedure that captures the correlation structure of the network well, using very few state variables. The algorithm begins assigning each particle a velocity v^d_i and a time $t_o=l_o/v_i$. The length l_o is a randomly generated advection length, the distribution of which is that of the network. The particle retains the velocity v_i until $t > t_o$, at which time a new random velocity is computed as $v_i = v^m_i + ?_i$ and a new random l_o is computed. Thus the statistical "state" is given by $?_i$ and t_o. This algorithm was implemented in a simple research code HD2T that runs on the AHPCRC CM5. The HD2T model produces plumes consistent with the SNC network model without performing a detailed high-resolution flow analysis. In the network model, velocity statistics are generated from the highly heterogeneous distribution of material properties. In the particle model the statistics are generated randomly by the algorithm described above.

An interesting aspect of the particle tracking technique is the ability to model length scales much smaller than are obtainable with direct simulation by PARFLOW or SNC. The dispersion rate of a direct simulation code is limited by the smallest spatial element in the grid. In particle tracking, the dispersion rate is limited numerically by the smallest step taken by a particle. In principle, the particle tracking method can capture dispersion at the smallest scale. However, to achieve such results, the numerical time step must be small. In dispersive transport problems, the time and length scales are related. To capture small dispersivity, a fine grid with an appropriately small time step is needed. For a particle tracking method, the spatial element of dispersion is no longer tied to the grid, and the time step, alone, is the limiting factor.

An advantage of the particle approach in implementation of the time-dependent dispersion is that it is easily adopted into finite element/finite volume engineering codes such as those routinely used at present in field-scale cleanup. Particle tracing offers particular advantage for parallel processing because each particle is independent. The independence of the particle motions simplifies implementation for the random walk routines. Inter-processor communication comes about only when computing mean flow velocities and data collation for post processing graphics, or calculating concentrations for chemistry computations.

Thus, a traditional subsurface model combined with the computationally efficient particle-tracking model can replicate the physical realism of the network model with much less computational burden. The important feature of the random walk model is the time-correlated velocity statistics created by heterogeneities. These statistics do not need to be obtained from a single high-resolution simulation that spans all scales, but can be assembled from several simulations that collectively span the scales. The current HPC Challenge Project combines data from the network, pore-scale Lattice-Boltzmann, and continuum-based finite element models to quantify velocity statistics of over a very large range of scales. ***These statistics are the basis of a procedure that permits relatively lower resolution flow computations, consisting of a few hundred thousand elements, to replicate the transport behavior of high-resolution analyses that contain billions of elements.***

3.1 Example of Leverage

Howington et al. (1997) describe an effort to use the SNC to predict the transport of a conservative tracer in a laboratory model experiment. The computer resources available at that time were insufficient to resolve the medium sufficiently. As a result the simulated plume was too dispersive such that peak concentrations at downstream sampling point were underestimated by nearly fifty percent. Resources made available through the Challenge project provide the computational power needed for the problem. To properly resolve this medium, a network with about 15 million network throats is needed. This requires approximately 8 GB of memory on 64 processors of the Cray T3E. By contrast, HD2T required only a trivial flow computation for the uniform flow field and simulated transport using 1,000,000 particles. Approximately 3000 CPU-minutes were used for the run. The breakthrough data for the HD2T simulation is shown in Figure 3, where it is seen that the peak concentration at the two sampling ports is quite good. The "tail" of the plume is not captured entirely, as indicated by the sharpness of the simulated plume versus that of the experiment. However, as the statistical distribution for velocity is improved, this detail should be likewise replicated well. As results from the very-high resolution SNC simulation become available, statistics of velocity can be approximated better.

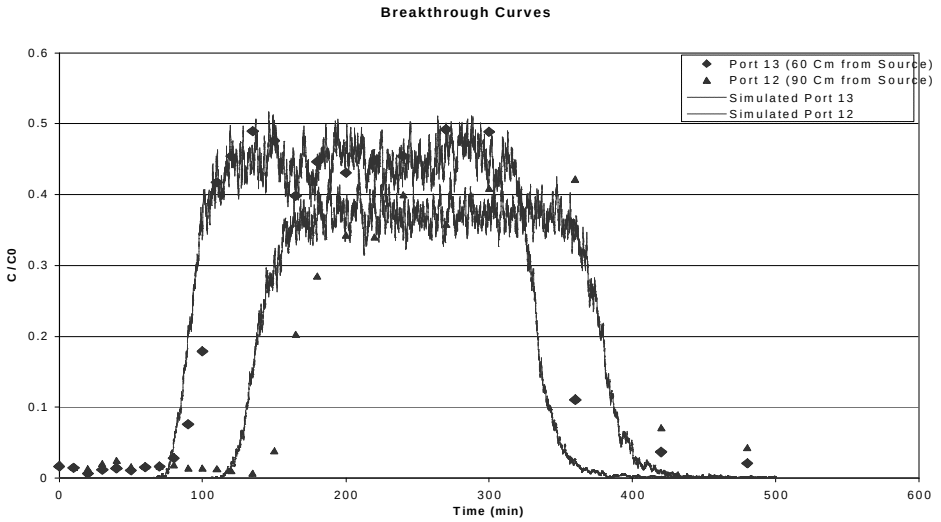


Figure 3. Comparison between experimental data and HD2T simulation**.**

3.2 Concentration Content

A fundamental concept of contaminant transport is concentration, which is the mass of constituent per unit volume of fluid. Implicit in this definition is that the mass is uniformly distributed throughout the volume. This definition presents an immediate problem to particle methods because it is unclear what reference volume should be used. The mass defined by a particle can be distributed to the sampling grid via a smoothing function. However, the concentration field does not converge as the grid is resolved. Rather, the concentration becomes infinite at the location of each particle as the grid cell size tends to zero. In essence, the particle tracking method is not well posed from a numerical standpoint unless the sampling grid size is somehow determined a priori as a "property" of the problem. For example, the grid size could be chosen such that the initial particle configuration correctly portrays the initial concentration field.

The same problem does not occur for the network because the mass is transported within volumes that are defined by the mixing rule for the network connections. The particle tracking method could similarly consider both mass and volume in the transport scheme. However, if the particles represent fixed volumes, the *maximum* concentration never changes. It is this maximum concentration that controls chemistry computation at the local level. The concentration based on a sampling grid merely represents the density of sub-volumes with each sub-volume having the maximum value. Although this is unrealistic, particularly for mature plumes, it contains an element of truth. The low concentration shown near the margins of a large plume may represent a sparse distribution of contaminant that has relatively high concentrations where it does occur. The network model and highly resolved continuum analyses captures this relatively well for, within a sampling grid cell, the distribution of concentration can be quite large. The use of the a priori sampling grid size in particle methods would miss this important feature of transport in porous media. Similarly, traditional transport equations appear to miss this feature entirely, even though they are well posed in the sense of the preceding discussion on particle methods. High-resolution schemes do not escape the problem entirely because of limits in available resolution. All practical models for contaminant transport tend to over-dilute the contaminants.

Proper coupling of processes does not require full spatial resolution of the medium. It does require a statistical depiction of the state at the micro-scale that includes more that just the mean concentration as is given by current methods. Referring to the particle method, each particle could have an initial volume, or spectrum of volumes, which would evolve as the particle moved. Such a law must be determined from high-resolution pore-scale computations. For example, in Figure 4 are statistics of concentration from an SNC simulation as a plume passes a selected observation point. The distribution of

concentration becomes broader as the plume travels. It is clear that the contaminant concentration seen by the solid phase varies greatly from the mean value. The goal of the current investigation is to derive a systematic evolution law from these statistics. The evolution law would be of the form of a history functional by which the probability distribution of concentration would depend on the history of mean concentration predicted from large-scale transport computations. In particular, the evolution law must account for hysteresis in the concentration distribution. The hysteresis is readily seen by comparing the distributions corresponding to approximately equal mean concentration but with different histories (e.g. the first and third distributions in Figure 4). As the mean concentration rises, the concentration is relatively broad but becomes more peaked. As the concentration falls, the distribution begins to flatten but with a noticeable lag in time. Isolated volumes of high concentration material remains even after the mean has fallen to low values. Details of the evolution law will be developed that will give the particle simulation the same concentration statistics as the SNC. By combining the SNC simulations with pore-scale simulations described in the next section, the statistical description should be valid at the smallest scale.

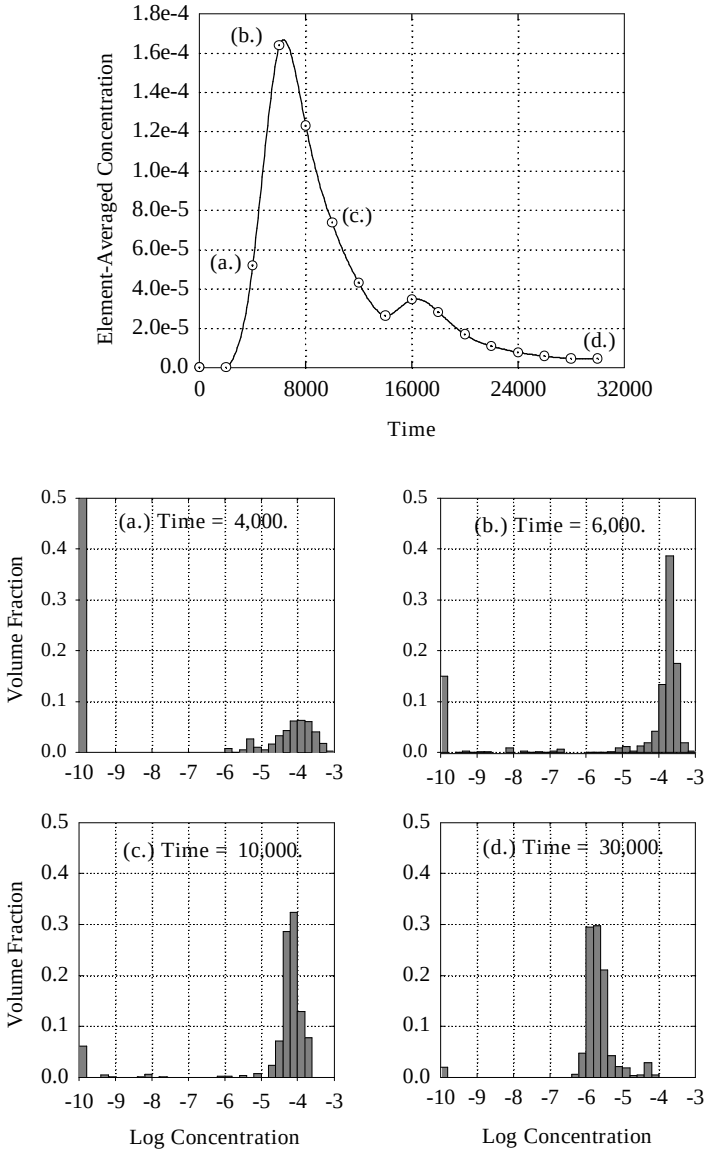


Figure 4. Breakthrough curve and concentration distributions at a fixed location downstream of an instantaneous injection of contaminant.

4. Stretching the Pore Scale

Both the SNC and PARFLOW are abstractions of the physical processes that occur at a smaller scale. These codes are useful for determining laws governing the statistics of dispersive transport and the evolution of concentration statistics but are more limited in dealing with coupled processes. Chemical and biological processes operate at the pore scale. While both SNC and PARFLOW simulations can be performed in which their fundamental elements are pore-scale size, these codes do not represent pore-scale behavior. Both are continuum porous media codes. PARFLOW is a direct discrete approximation to the continuum mechanics equations for porous media whereas the SNC is a discrete analogy with behavior that converges to the continuum ideal. In the case of both, pore scale behavior is represented statistically and is therefore valid only when "observed" from larger scales. To derive appropriate laws for processes that are coupled at the pore scale, a true pore-scale representation is needed. A number of such simulation models have been used in academic research but are limited in size and scope. What is needed for the present research is a pore-scale model that spans between the level at which fluid flow within the pores is governed by the Navier-Stokes equations for a domain size that represents a REV. Otherwise, boundary conditions dominate the simulations or larger-scale structures do not develop. It is these larger scales that the SNC and PARFLOW can model.

Even with the "reduced" requirement of only reaching the next scale level, currently available HPC resources are marginally adequate to meet the project needs. Consider the "worst case" of pore-scale measurements using Latice-Boltzman models. Pore-scale models simulate the movement of fluid within the detailed geometry of a porous medium. Void and solid spaces within a soil, rock, or sand sample are rendered in a digital image, and the equations of flow and transport are solved, using the digital image to provide detailed boundary information. Although pore-scale models are the most accurate representation of a porous medium, they are limited by computer capacity in the maximum size of the simulated physical domain. Computer memory must be allocated to resolve the pore spaces with sufficient resolution to approximate the flow, whether a regular or irregular mesh is used. The domain size must allow the plume to grow large enough that it can serve as the initial distribution for a network model. Continuity between the two models is achieved if subsequent evolution of the plume is similar in the two models. Such a pore-scale model would have linear dimensions of several hundred particles. The smallest element in this domain is dictated by the accuracy of the flow and transport within the pore space. A sufficient number of lattice cells must exist within the pore space resolve circulation fields at the pore level. Considering a minimal requirement to be 10 lattice points for the nominal particle size, LB simulations with approximately 1 billion lattice sites are needed.

4.1 Role of Pore-Scale Simulation

The goal of pore-scale simulation is to develop a "virtual" laboratory for the analysis of flow through porous media and comparisons with physical (laboratory-scale) experiments. Research objectives include (a) the study of velocity distributions in mono-disperse and bi-disperse colloidal sphere packs, and (b) the study of conservative solute particle displacement distributions (dispersion) in random porous media. There are two high-performance computing (HPC) objectives that support the research objectives. The first is to develop a fluid-dynamics code that can be applied routinely, with low human overhead, to the simulation of flow through porous media. This class of flow problems is typified by low Reynolds numbers, $R_e = v^m d / ? < 100$ (where d is the characteristic soil particle diameter) and by irregular pore geometries.

The second supporting HPC objective is to develop a transport simulation code which can utilize flow fields generated by the fluid-dynamics code to predict dispersion in the high Peclet-number regime, where typically $P_e = v^m d/D^m > 100$. The combined research and HPC objectives represent the development of an enabling technology for studies of pore surface reactivity, and are therefore significant to a wide range of subsequent investigations. The objectives are significant from a computational perspective because current studies in the physics of pore-scale flow rely upon coarse-grained simulations, and have not established grid convergence.

4.2 Methodology

The specific numerical methods selected for computer implementation in accord with the HPC objectives include: the lattice Boltzmann (LB) method, for simulating low Reynolds-number flows; and a stochastic particle-tracking method for simulating solute transport in the high Peclet-number regime. The choice of CFD algorithm was guided initially by the concept of a 3-D digitized image of a porous medium, obtained by tomographic or spectroscopic techniques, or synthesized computationally on the basis of desired statistical properties. This concept suggested a one-to-one mapping of image voxels to computational cells or elements. An equally important consideration was the low Reynolds-number regime of porous media flows, which permits the use of basic numerical methods, such as simple upstream weighting in a finite-difference formulation.

The lattice Boltzmann (LB) method approaches fluid dynamics from the microscopic, kinetic level by solving the discrete Boltzmann equation, which was originally formulated as a description of dilute gas behavior. The capacity of the discrete Boltzmann equation for recovering Navier-Stokes dynamics for *liquids* has been established, both in theory and practice, for single-phase flows in the low Mach-number regime (Chen et al., 1991, Maier et al., 1997, 1998). However, the feature which makes these methods attractive for pore-scale simulation is a computer implementation with very low memory requirements—about 10 floating point words per computational node, including the macroscopic state variables (density and three velocity components). Of course soil and rock porosity range from below .20 (sandstone) to over .40 (sand), suggesting that better memory use could be obtained by an unstructured scheme. Unfortunately, the net reduction in memory requirements is only on the order of 50 percent to 70 percent because of the increased storage requirement for pointers. In a parallel environment, the communications overhead makes the use of pointers even less attractive. So, to minimize initial software development time, we employed a regular grid formulation of the LB method, at the expense of approximately a factor of two in memory requirements.

The LB code uses an explicit, finite-difference formulation on a fixed grid to achieve low memory overhead per computational node. The numerical properties and performance have been validated carefully, and the code has been used routinely in the production of research results. The first implementation was developed in the data-parallel programming model, for the Thinking Machines CM-5 massively parallel computer. A new implementation was prepared in FY97 using the message-passing model, with resulting improvements in memory utilization and performance. The initial development plan for a solute transport algorithm called for an extension of the LB method, one which would recover the dynamics of the advection-diffusion equation, and at the same time achieve algorithmic and software compatibility with the LB CFD code. While previous theoretical and applied numerical research by others had indicated the feasibility of this approach, our own numerical testing revealed overly-diffusive behavior in simulating high Peclet numbers in the same pore geometries used for CFD simulations. This deficiency suggested the need for a more sophisticated numerical method for handling sharp concentration gradients. However, physical experiments employed NMR spectroscopy techniques that would require a Fourier analysis of simulated concentration distributions into a form that is comparable to measured molecular displacement distributions. This additional complexity favored a Lagrangian approach to the description of transport phenomena.

An Eulerian-Lagrangian method was ultimately selected to simulate solute transport in the production code. The velocity field for the fluid is obtained from the LB CFD code, which gives the drift velocity at fixed lattice points in the digital pore image. An ensemble of discrete particles, whose trajectories are computed by solving a Langevin equation, represents the solute. The Lagrangian approach was selected for ease in simulating the high Peclet-number regime. Theory shows that the Lagrangian approach recovers the behavior of the classical advection-diffusion equation (Gelhar, 1993), without the particular numerical problems encountered when solving the advection-diffusion equation for high Peclet numbers. A drawback of the Lagrangian approach is the poor convergence rate with respect to the number of particles. The computer implementation of the Lagrangian approach uses dynamic linked lists to store particle coordinates, and message passing (MPI) is used to share coordinates which cross inter-processor boundaries.

4.3 Verification of Pore-Scale Model

Viscous fluid flow was simulated through a variety of laboratory-scale packed beds, including columns of glass beads and periodic random bead packs. While a digital image of a column of beads corresponds closely to the physical laboratory system, periodic random packing allows the simulation of arbitrarily large domains. Periodic packs of spheres are created by a Monte-Carlo algorithm that relaxes a configuration of hard spheres in a manner similar to conventional molecular dynamics simulations. The relaxation algorithm imposes periodic boundary conditions so spheres on one side of the simulation box "feel" the spheres on the opposite side. The resulting packs have periodic continuity, and represent a unit "cell" in larger periodic array of such cells. Packed beds that represent a laboratory column and those that have periodic continuity share certain flow characteristics, stemming from the randomness of both types of packs. The most salient characteristic is that the normalized velocity distribution (the density function $P(v/v^m)$) of flow in mono-disperse random sphere packs collapse onto a single profile for a variety of Reynolds numbers, porosities, and permeabilities. Ongoing work with bi-disperse packs suggests this result may generalize to a broader class of porous media. One practical consequence of this finding is that a periodic system and a column pack can be used more or less interchangeably to accomplish simulation tasks.

With increasing spatial resolution, the simulated velocity distribution in a random pack converges to a velocity distribution with a sharp peak near zero (Figure 5). In fact, we found that spatial resolution has a greater influence on the shape of the velocity distribution than does the particular random packing arrangement. This grid convergence result was possible only through the use of HPC resources, and was critical to proper comparisons with velocity distributions from NMR spectroscopy. The solute transport code has been used to reproduce tracer displacement distributions from physical experiments, to establish a valid foundation for application of the code to other simulation problems. To date, we have verified that the code correctly reproduces the first moment of the tracer distribution.

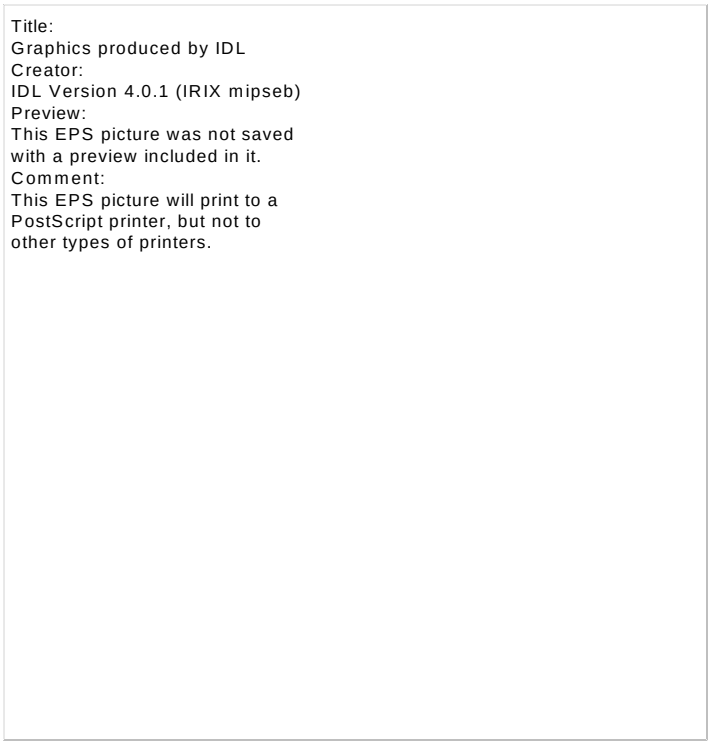


Figure 5. Velocity histograms from the simulation of flow through beadpacks with spatial resolution varying from *d = 9* to *73* lattice spaces per bead diameter.

We are presently working toward verification of the second moment of the simulated tracer distribution through comparisons with published physical experiments. The time derivative of the second moment, or dispersivity, characterizes the rate of solute spread, and is widely used to parameterize groundwater models. Simulated dispersivity in random packs does scale with Reynolds number according to experimental results, but we are still investigating why point estimates of dispersivity are uniformly below certain experimental results. It is not yet clear whether this discrepancy is a defect in the simulation or possibly an experimental artifact. We have, however, observed a strong relationship between the magnitude of the dispersivity estimate and the heterogeneity of the packing. Even in the mono-disperse case, layering of beads can significantly change the short-term spreading behavior.

4.4 Computational Performance

The LB code was originally developed in a HPF-style programming model for the CM-5. The code scaled well on that architecture, but obtained little more than 10 per cent of the peak performance, primarily because of a high ratio of memory-reference to computation and a high ratio of inter-processor communication to local computation. Scaling data is given in the following tabulation:

Lattice	Processors	Memory (Gb)	CPU (min)	Disk (Gb)
128^3	64	0.2	15	0.08
256^3	256	1.4	33	0.67
512^3	512	9.2	113	5.37

The code was migrated from the AHPCRC CM-5 to the CEWES IBM-SP. The code was rewritten using the Fortran 77/MPI programming model, with resulting improvements in processor and memory utilization on the IBM-SP. Initial experience with the IBM indicated a 12-fold improvement in speed on a per-processor basis. Performance measures were collected for some typical production runs. The problem simulated was a periodic random pack consisting of 1,000 mono-disperse beads that was represented on a 512 x 512 x 512 lattice. Densities and momenta were calculated for a range of Reynolds numbers. Timings were collected after 4,000 LB time steps (the number of steps computed on the CM-5 before user check-pointing and normal job termination). The production job required 2.35 hours on 128 processors of the IBM-SP, compared to 6.83 hours on 512 processors of the CM-5. A conservative estimate of performance for the IBM-SP is 188 MFLOPS per processor, or approximately 35 per cent of peak. Four similar production runs were made on the IBM, using 64, 96, 128, and 192 processors. Each of the four runs performed the same amount of fluid dynamics calculation, but those runs used more processors and performed slightly more inter-processor communication. A plot of speedup (Figure 6) shows the rate of decrease in overall efficiency, as the inter-processor communication becomes more significant. In addition, a production test run utilizing one billion computational nodes on 216 processors of the CEWES IBM-SP was completed as technical challenge that is essential for meeting project goals. The test involved the calculation of 1,000 LB time steps on a 1024 x 1024 x 1024 grid, modeling pore-scale, viscous flow through periodic random packs. The test required 3.9 hours of wall time. It is estimated that 12,000 time steps are needed to reach steady state, a figure that clearly points to the need for extensive HPC resources for this investigation.

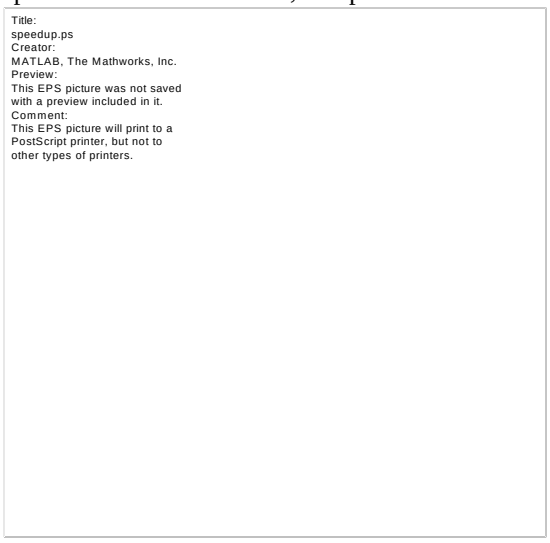


Figure 6. Speedup curve for the LB model on the IBM SP.

5. Future Work

In the coming year efforts will be focused on dispersion of immiscible fluids. The requirements for the pore-scale computation of immiscible are on the same order as those presented above for miscible transport. Preliminary computations indicate that SNC simulations for immiscible fluids require an order of magnitude increase in CPU time because additional iterations for non-linear behavior are required. The particle tracking technique will be elevated to a practical level by coupling to a non-uniform flow field, incorporating chemistry by coupling to a concentration field, and multiple-phase tracking. Ultimately, the

particle tracking computations will require the same order of HPC resources now required for the high resolution simulations, but with greater richness in detail of physical process than is possible with the current generation of computer codes.

6. Summary

The U.S. Department of Defense is continuing research and development whose primary goal is development of a comprehensive groundwater modeling system for use in the accurate assessment of cleanup strategies for contaminated groundwater resources at military installations. Much of the system is already in place or will come about through a systematic extension to existing technology. A critical requirement is the development of numerical subsurface models that properly bridge the range of scales. This requirement is the key to accurate modeling of flow, transport, and remediation in heterogeneous porous media. HPC plays a central role in the development of these next-generation models by providing data that are virtually impossible to obtain by experimental means. Substantial improvements in the effectiveness and defensibility of remedial actions are expected through implementation of this technology, thereby resulting in significant savings in remediation costs.

Application of this technology is a concerted effort using computational models at various scales. The LB model provides the velocity statistics for the small-scale components (spatial and temporal) in the PARFLOW and SNC simulations. The computational "leverage" provided by the particle method is encouraging, because it provides the means to apply the scientific principles established from high-resolution multi-scale computations to practical design of remediation schemes.

Acknowledgements

The paper was prepared from research and development conducted under the Groundwater Modeling Program of the U.S. Army Engineer Waterways Experiment Station, Vicksburg, MS. This work has been supported by the Army Environmental Center. Permission was granted by the Chief of Engineers to publish this information.

References

Chen, S., Diemer, K, Doolen, G, Eggert, K, Fu, C., Gutman, S., and Travis, B., "Lattice Gas Automata For Flow Through Porous Media," *Physica D*, Vol 47, pp 72-84, 1991.

Cushman, J. H. and B. X. Hu, "A Resume of Nonlocal Transport Theories," Stochastic Hydrology and Hydraulics, Vol 9, pp 105-116, 1995.

Gelhar, L. W., *Stochastic Subsurface Hydrology*. Prentice-Hall, Englewood Cliffs, NJ, 1993.

Holland, J. P. "Department of Defense Groundwater Modeling Program: An Overview," in *Subsurface Fluid-Flow Modeling*, ASTM STP 288, edited by J. D. Ritchey and J. O. Rumbaugh, American Society for Testing and Materials, Philadelphia, PA, 1996.

Horner, D. A. "Application of DEM to Micro-Mechanical Theory for Large Deformation of Granular Media," Ph.D. Dissertation, University of Michigan, 1997.

Horner, D. A., Peters, J. F., Howington, S. E., and Hryciw, R., "Effects of Grain Size Distribution on the Physics of Particulate Materials," in *Proceedings of the Institute for Mechanics and Statistical Physics of Particulate Media Workshop*, Institute for Mechanics and Materials, University of California, San Diego, pp 86-89, 1994.

Horner, D. A., Carrillo, A. R., Peters, J. F., and West, John E., "Application of DEM to Micro-Mechanical Theory for Large Deformations of Granular Media," *Mechanics of Structures and Machines*, Accepted for publication.

Howington, S. E., Peters, J. F., and Illangasekare, T. H., "Discrete Network Modeling for Field-Scale Flow and Transport Through Porous Media", Technical Report CHL-97-21, September 1997.

Hryciw, R., Raschke, S. A., Ghalib, A. M. Horner, D. A., and Peters, J. F., "Video Tracking for Experimental Validation of Discrete Element Simulations of Large Discontinuous Deformations," *Computers and Geotechnics*, Elsevier Science Ltd., 17 pages, 1997.

Koch, D. L., and Brady, J. F., "A Non-Local Description of Advection-Diffusion with Application to Dispersion in Porous Media," *Journal of Fluid Mechanics*, Vol. 180, pp 387-403, 1987.

Maier, R. S., Kutsovsky, Y., Nivarthi, S., Davis, H. T., Grunau, D. W., Howington, S. E., and Bernard, R. S., "Pore-Scale Flow and Transport Calculations Using the Lattice Boltzmann Method," in *Next-Generation Environmental Models and Computational Methods*, G. Delic and M. F. Wheeler, editors, pp 259-263, SIAM, Philadelphia, 1997.

Maier, R. S., Kroll, D. M., Kutsovsky, Y. E., Davis, H. T., and Bernard, R. S., "Simulation of Flow Through Bead Packs Using the Lattice Boltzmann Method," *Physics of Fluids*, 1998.

Neuman, S. P., "Eulerian-Lagrangian Theory of Transport in Space-Time Nonstationary Velocity Fields: Exact Nonlocal Formalism by Conditional Moments and Weak Approximation," *Water Resources Research*, Vol 29, No 3, pp 633-645, 1993.

Peters, J. F. and Howington, S. E., "Pre-Asymptotic Transport Through Porous Media," in *Next Generation Environmental Models: Computational Methods*, G. Delic and M. F. Wheeler, editors, pp 271-280, SIAM, Philadelphia, 1997.

Poeter, E. P., W. L. Wingle, and E. V. Edris, "Utilizing Disparate Types of Data at a Complex Site to Reduce Uncertainty Associated with Estimated Parameters," in *Proceedings of 27th Congress of International Association for Hydraulic Research, Theme C Groundwater: An Endangered Resource*, pp 343-348, 1997.

Scheibe, T. D. and Cole, C. R., "Non-Gaussian Particle Tracking: Application to Scaling of Transport Processes in Heterogeneous Porous Media," *Water Resources Research*, Vol 30, No 7, pp 2027-2039, 1994.

Schmidt, J. S., "3-D Adaptive Grid Modeling for Groundwater Mechanics,"' in *Next Generation Environmental Models: Computational Methods*, G. Delic and M. F. Wheeler, editors, pp 265-270. SIAM, Philadelphia, 1997.

Tompson, A. F. B. and Gelhar, L. W., "Numerical Simulation of Solute Transport in Three-Dimensional, Randomly Heterogeneous Porous Media," *Water Resources Research*, Vol 26, No. 10, pp 2541-2562, 1990.

Tompson, A. F. B., Rosenberg, N. D., Bosl, W. J., Falgout, R. D., Smith, S. G., Shumaker, D. E.; and S. F. Ashby, "On the Use of High-Performance Simulation in the Management of Groundwater Resources in Large Aquifer Systems," in *XXVII IAHR Congress - Groundwater: An Environmental Resource*, pp 337-342, 1997.

Key Words

Groundwater, Scale Effects, Heterogeneity, Virtual Laboratory